Towards Improving the Availability of Service Compositions

Dominik Dahlem, Lotte Nickel, Jan Sacha, Bartosz Biskupski, Jim Dowling, and René Meier
Distributed Systems Group
Department of Computer Science
Trinity College Dublin
{dominik.dahlem,nickell,jsacha,biskupski,jim.dowling,rmeier}@cs.tcd.ie

Abstract— The proliferation of open Internet-scale service-oriented platforms based on standards, such as WSDL, SOAP and BPEL, enables the composition of independent web services into new value-added services. Such service compositions define the information flows between autonomous and potentially heterogeneous services across the boundaries of independent provider organisations. The availability of individual services in such Digital Ecosystems is likely to be variable due to fluctuating usage load and resource limitations imposed by a service provider's infrastructure. This problem becomes more acute as the number of services in a composition increases. This paper presents a mediation model for improving the availability of composed services. The mediation model masks failures in a service composition by transparently selecting (and executing) an alternative composition at runtime. Service consumers use a common interface to a set of functionally equivalent service compositions while a selection mechanism identifies the most suitable (alternative) service composition. An evaluation of our implementation of the proposed mediation model demonstrates that the consumer perceived availability of value-added services can be improved significantly.

Index Terms— Service-Oriented Architectures, Service Composition, Dynamic Service Selection

I. INTRODUCTION

Web services support application-to-application communication over the Internet. These services provide standardised interfaces and are ignorant towards protocols in order to allow services from a variety of providers to interact. Several approaches to Web service composition have been proposed to facilitate the combination of a set of Web services into a more complex business logic. These service compositions accept invocations from consumers, coordinate the invocations of the underlying Web services, and return the results to the consumers.

Web service composition architectures, however, suffer from two major problems: they depend on the availability of the individual Web services and their correct behaviour, and they lack the flexibility to replace a failed Web service with a redundant alternative. Generally, the availability of a basic service cannot be guaranteed and therefore requires each service provider to tackle this problem individually. If one of these services is not available or fails during execution, a dispatcher component needs to be able to dynamically switch to alternative Web Services that provide equivalent functionality in order to fulfil the consumer request. These problems

escalate with the introduction of service compositions. The replacement of a service composition with an alternative composition increases the functional complexity involved and also increases the response time of an invocation.

This paper proposes a mediation model for improving the availability of value-added services based on service compositions. The workflow mediation model offers service instance and service composition failure resilience through the ability of coordinating between alternative service compositions. Internet-scale Digital Ecosystems can take advantage of this architecture in order to propel business interactions between small and medium-sized enterprises (SMEs). More importantly, the growing role of smaller firms emphasises the importance of improving the availability of digital services, because they might not have an expensive infrastructure in place that guarantees high-availability. Consequently, consumers and service providers benefit, because an improved reliability model to expose services improves the consumer perception and encourages consumers to use this service again in the future.

II. MOTIVATION AND RELATED WORK

Services deployed in an open service-oriented architecture are inherently autonomous. Consequently, value-added composite services are vulnerable, because failures immediately propagate up the service chain to the consumer, if no compensating actions are explicitly or implicitly defined. Recent research activities concentrated on improving the service infrastructure by removing single points of failure through the introduction of P2P architectures [1], [2] or by providing dynamic selection mechanisms based on consumer preferences [3], [4], [5], [6], [7].

Static composition describes the process of designing a workflow based on well-known existing services. BPEL has become one of the accepted standard languages on top of WSDL that defines a process-centric model for the formal specification of the behaviour of business processes based on the interaction of the executable process and its partners [8]. The BPEL process itself is unaware of the underlying dependency on the concrete WSDL bindings and service endpoints and thus allows for some dynamic assignment of service endpoints during the execution of the BPEL process [9]. However, some BPEL-based workflow management systems

do not support the internal assignment or selection of service endpoints and therefore require the service developer to define the actual endpoints of the services before the composition can be deployed [1].

Often, static workflows are too rigid, because they cannot cope with a frequently changing environment. Motivated by the requirement of more dynamic service-oriented architectures, eFlow [3] decouples the service selection from the process definition in order to allow the selection process to be governed by consumer constraints, such as low-cost. Consequently, the best currently available service can be selected at runtime. The dynamic selection rule requires the management of conversations with the dynamically selected services. eFlow tackles this problem by maintaining a repository of conversations which are associated with at least one service. If two or more services share the same interface then the repository keeps one conversation representing those services, whereas functionally similar services require individual conversations. When eFlow selects a service it also selects the associated conversation. Li et. al. [5] present a similar model of dynamic service composition. Instead of relying on a repository of conversations, they use adaptable connectors to reconfigure the candidate services and service selection function.

While Li et. al. [5] and Casati & Shan [3] allow dynamic selection of services based on some consumer preferences, their work does not align with middleware that incorporates an explicit Quality of Service (QoS) model. [4], [6], [7] reconcile QoS models with service-oriented architectures, in particular to be able to improve the consumer perception of a service by selecting the best one available at a given time.

Other research focuses on eliminating single points of failure by leveraging P2P systems in order to provide a decentralised service broker architecture [1] or to distribute the execution of a workflow [2].

Previous work [1] introduced a binding- and port-agnostic service composition framework that extends an existing open-source BPEL workflow management system and leverages a P2P broker architecture in order to bind to services at runtime. In this architecture, services are referenced by a logical identifier which is then mapped to the actual endpoint at runtime. The P2P system provides the binding- and port information encapsulated in a service proxy object which can then be invoked by the consumer. In contrast to our approach, Shen et. al. [2] reported an approach that orchestrates and executes workflows in a distributed fashion. Their decentralised workflow system self-manages into service groups with the same capabilities. When a peer accepts and executes an assigned task, it will discover the community with the required capability for the succeeding task. The community elects the best member of this community to accomplish this task. As a result, the execution path of a business process is distributed among a set of peers and promises better scalability and performance of the whole system.

The presented work on dynamic service composition models leverages consumer constraints in order to dynamically select the best performing service. However, they do not explicitly deal with improving the availability of services over time. We argue that existing work can be extended to provide more available services for a digital ecosystem by applying our mediation model. The mediation model can be understood as a pattern that may complement existing dynamic service selection mechanisms and improve service composition deployments.

III. THE MEDIATION MODEL

A. Architecture

Our hierarchical mediation model is based on a multi-layered architecture with distinct functional requirements for each layer. Figure 1 delineates the elements of the model, which encompasses three layers.
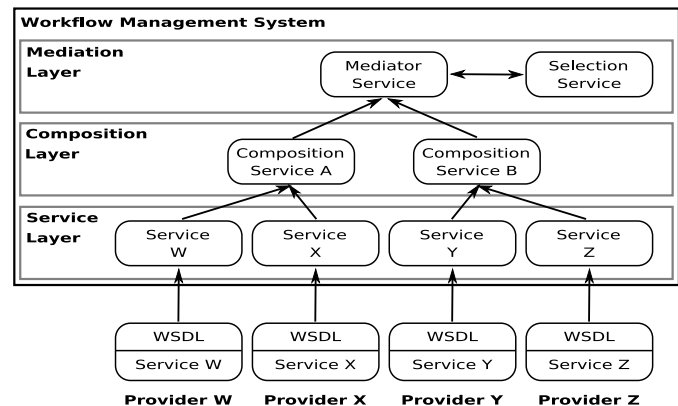

Fig. 1. The Multi-layered Service Composition Model

The Service Layer provides a set of candidate services for the Composition Layer. A candidate service is a Web Service described in WSDL that represents a remote service. At design time only the abstract definition of the interface is of interest to the Composition Layer and is therefore independent towards the underlying interaction paradigm of the candidate service. In fact, it is also independent to the endpoint binding. If multiple services implement the same interface, the workflow management system is more resilient to service instance failure.

The Composition Layer defines at least one workflow that coordinates the execution of one or more candidate services. The workflows on this layer define semantically the same business logic. Internally, however, the business logic of the workflows may differ in the way they manage the control and data flow among the candidate services. For example one workflow is likely to use a different set of service providers and possibly requires a higher degree of service interaction than others.

Finally, the Mediation Layer exposes a single interface to the consumers of a composite service while allowing the dynamic selection of the most suitable workflow at a given time, which is provided by the Selection Service. Essentially, the Mediator Service is a composite service that coordinates among the workflows provided by the Composition Layer and the Selection Service. In contrast to the Mediator Service, however, the workflows defined in the Composition Layer are not exposed to the consumers. Instead the Composition Layer represents a set of possible conversations with remote services

with similar functionality. The selection algorithms chooses among the registered workflows in the Composition Layer and returns it to the Mediator Service. If the invocation of the selected workflow fails, then the Mediator Service masks the failure and delegates the request to the next workflow in the Composition Layer until a threshold `maxTries` is reached.

IV. IMPLEMENTATION

The mediation model has been implemented in a sample online travel agent. The service compositions are defined in the BPEL language and deployed each in the ActiveBPEL engine.
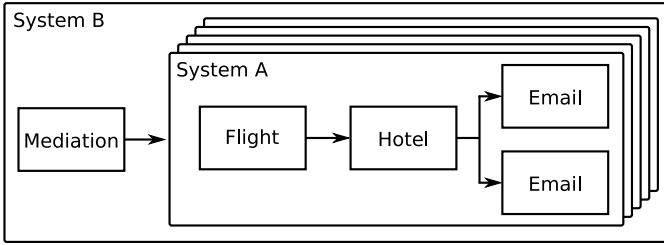

Fig. 2.   Online Travel Agent Scenario

Figure 2 provides an overview of the online travel agent service. Five service compositions (System A) were developed that implement the core functionality of booking flights and hotels and receiving a status email. While each of these compositions adhere to the same structural behaviour, they internally use different service providers to demonstrate a true alternative composition selection. Moreover, a mediation service (System B) was developed that selects dynamically among the core service compositions. In this paper a simple selection service is considered, whose selection mechanism initially assigns equal probabilities to all registered service compositions. If the invocation of a specific service composition yields a failure, the selection service assigns a zero probability to the execution of the same service composition thereby allowing for the selection of an alternative composition. In the worst case, the selection of service compositions is exhausted and an error will be returned to the consumer. The mediator service maintains this state for the life time of a session, i.e., the interaction between receiving a request, processing it, and finally sending a response back to the consumer. Workflow management systems assure that parallel sessions from different consumers can co-exist without conflicts.

V. EVALUATION

The mediation model is evaluated based on a travel agent scenario and it is shown that our proposed model shows significant improvements over a simple service scenario. In particular, the reliability of System A in isolation is compared to the reliability of System B with five redundant service compositions. While system A describes a basic booking composite service using a flight booking, hotel booking, and a redundant email service, system B is an implementation of our mediation model that transparently dispatches consumer requests between five workflows that adhere to the structure of

System A, but use different service providers. For the purpose of this evaluation the failure rate of the underlying basic services was fixed at 5% and 20% to demonstrate the impact of those Service Layer failures on both, System A and System B. The failure rate corresponds to a simple failure metric without considering service downtime. Thus, the failures are considered independent random variables for the purpose of this evaluation.

$$
\begin{aligned}
r_c(p_s) &= [(1-p_s)*(1-p_s)]*[1-(p_s*p_s)] \quad (1) \\
r_m(p_c) &= p_c^5 \quad (2)
\end{aligned}
$$

Equation 1 and Equation 2 present the reliability functions of the service compositions in the Composition Layer and the mediator service in the Mediation Layer respectively, where p_s is the failure rate of the individual services in the Service Layer and p_c is the failure rate of the service compositions with a reliability of 90% and 61% for the failure rates 5% and 20% respectively. The service compositions are based on the same structure function, therefore the structure function of the mediator service, $r_m()$, can be reduced to p_c^5 with a reliability of >99% and 99% for the failure rates 5% and 20% respectively.

The mediated model (System B) thereby shows significant higher reliability levels compared to the basic model (System A) in isolation (with an increase of approximately 10 percentage points at the 5% and 38 percentage points at 20% Service Layer failure rate respectively). In particular, our research hypothesis is that the reliability level of System B is greater than the one of System A, $p_B > p_A$ for all observed Service Layer failure rates, whereby the magnitude of the difference in the failure rates increases with an increase in the Service Layer failure rate. Analogue, the null hypothesis states no difference among System A and System B or higher reliability levels for System A.

The mediation model is evaluated with independent tests for each Service Layer failure rate. First, a one-tailed power analysis was conducted with the parameters power of 0.95, significance-level of 0.05, and an effect-size based on the theoretical difference in reliability between System A and System B for each Service Layer failure rate to estimate the sample size that is required to yield significant results. Then a sample is drawn for each failure rate out of a population of 12500 test runs in total. During the test runs a boolean variable was collected indicating whether the invocation to either System A or System B was successful. A failure in System A indicates a failure of at least one flight or hotel booking or both email services, while a failure in System B indicates that all service compositions did not complete successfully.

Table I summarises the results of the power analysis and the sampled data. Note that all the theoretical reliability levels are within the reported second-order corrected confidence intervals of the samples. Also, with increasing Service Layer failure rates the sample size decreases. This can be explained by the fact that the gap between the reliability levels of both systems increases dramatically and consequently fewer samples are

TABLE I

EXPECTED AND SAMPLED RELIABILITY LEVELS

Expected Reliability (P), sample size (N), and the 95% confidence
interval (CI) of all Service Layer failure rates

	5%			20%		
	P	N	95% CI	P	N	95% CI
System A	0.9	101	[0.88; 0.98]	0.61	22	[0.34; 0.74]
System B	>0.99	101	[0.98; 1]	0.99	22	[0.90; 1]

necessary for the test to show significant results with the given parameters for the power-test as explained above.

To complete the tests a Fisher's exact test was conducted for each Service Layer failure rate to determine if there are any non-random associations between the sampled data of both, System A and System B. The Fisher's exact test was used instead of the Chi-Squared test, because some of the cells in the matrix shown in Table II have a count of less than 10 and zero cells.

TABLE II

COMBINED CONTINGENCY MATRIX AND RESULTS FROM THE FISHER'S
EXACT TESTS

	5%		20%	
	Success	Failure	Success	Failure
System A	95	6	12	10
System B	101	0	22	0
Fisher's Exact Test				
p-Value	0.001		<0.000	
95% CI	[0; 0.63]		[0; 0.23]	

Computing the p-values with the Fisher's exact test reveals a statistically significant difference between the reliability levels of System A and System B, because all the p-values are below the fixed 0.05 significance mark (see Table II for the exact results). Additionally, the 95% confidence interval is supportive of this fact.

Therefore, the null hypothesis can be rejected in favour of the research hypothesis being true. Consequently, the mediation model shows significantly higher reliability levels than the basic booking workflows.

VI. DISCUSSION AND CONCLUSION

An experimental study of an online travel agent's reliability shows a dramatic decrease with an increasing failure rate of the underlying services. With the proposed mediation model, however, the reliability significantly improves the situation by masking internal failures and dispatching to alternative service providers. Although the experimental study was based on a sample scenario it relates to any online business that is composed of third party services. Therefore, the improvement that the mediation model offers has substantive consequences, especially in a business that depends on services that are deployed in Internet-scale environments. Both consumers and service providers benefit, because a satisfied consumer most likely uses the same service again in the future. Moreover, generally no assumptions on the reliability of Internet services can be made. As a result service providers have to control

for services being unavailable in order to not suffer from unexpected failures from third party services.

Finally, the mediation model presented in this paper can be implemented with any composition language and consequently does not require changes in existing service infrastructures. Additionally, the Composition and Mediation Layers can be distributed between different hosts to further reduce the reliance on single points of failure.

In this paper we proposed a mediation model to mask internal failures of service compositions by dispatching to redundant alternatives. Our results showed that a mediated model presents significantly higher reliability levels. Our future work involves merging the approaches of preference-based service selection with our mediation model. Those activities involve more sophisticated probabilistic service selection methods then the one presented in this paper.

REFERENCES

[1] D. Dahlem, D. Mckitterick, L. Nickel, J. Dowling, B. Biskupski, and R. Meier, "Binding-and port-agnostic service composition using a p2p soa," in *Proceedings of the International Workshop on Dynamic Web Processes (DWP 2005)* (K. Verma, A. Sheth, M. Zaremba, and C. Bussler, eds.), pp. 61–72, IBM Research Division, December 2005.

[2] J. Shen, J. Yan, and Y. Yang, "Swindew-s: extending p2p workflow systems for adaptive composite web services," in *Australian Software Engineering Conference, 2006*, pp. 9–17, IEEE, April 2006.

[3] F. Casati and M.-C. Shan, "Dynamic and adaptive composition of e-services," *Information Systems*, vol. 26, pp. 143–163, May 2001.

[4] X. Gu and K. Nahrstedt, "Distributed multimedia service composition with statistical qos assurances," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 141–151, 2006.

[5] G. Li, Y. Han, Z. Zhao, J. Wang, and R. M. Wagner, "Facilitating dynamic service compositions by adaptable service connectors," *International Journal of Web Services Research*, vol. 3, pp. 68–84, January 2006.

[6] Y. Liu, A. H. Ngu, and L. Z. Zeng, "Qos computation and policing in dynamic web service selection," in *WWW Alt. '04: Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, (New York, NY, USA), pp. 66–73, ACM Press, 2004.

[7] D. A. Menasce, "Composing web services: A qos view," *IEEE Internet Comput.*, vol. 8, no. 6, pp. 88–90, 2004.

[8] N. Milanovic and M. Malek, "Current solutions for web service composition," *IEEE Internet Comput.*, vol. 8, pp. 51–59, November 2004.

[9] F. Curbera, R. Khalaf, N. Mukhi, S. Tai, and S. Weerawarana, "The next step in web services," *Communications of the ACM*, vol. 46, pp. 29–34, October 2003.